DOCUMENT RESUME

ED 310 383 CS 211 956

AUTHOR Lovejoy, Kim Brian; Lance, Donald M.
TITLE Discourse Analysis: Part I, Information Management and Cohesion.
PUB DATE Aug 88
NOTE 59p.; Preliminary version of this paper was presented at the International Systemic Congress (15th, East Lansing, MI, August 1988). A shorter version of this paper was presented at the Annual Meeting of the Conference on College Composition and Communication (40th, Seattle, WA, March 16-18, 1989). For part II, see CS 211 957.
PUB TYPE Reports - Research/Technical (143) -- Speeches/Conference Papers (150) -- Reports - Descriptive (141)

EDRS PRICE MF01/PC03 Plus Postage.
DESCRIPTORS *Cohesion (Written Composition); *Discourse Analysis; Intonation; Language Research; Models; Semantics; Syntax; Writing for Publication
IDENTIFIERS Descriptive Research; *Information Management (Written Composition); *Text Factors

ABSTRACT

Combining linguistics and composition studies, this paper (part 1 of a two-part article) proposes a model for the analysis of information management and cohesion in written discourse. It defines concepts of discourse analysis--specifically information management, syntax, semantic reference, lexicon, cohesion, and intonation, with examples taken from publications in psychology, biology, and history. The paper discusses key terms related to reference and information, including "information unit," "theme," "rheme," "topic," "comment," "given," "new," "focus," "markedness," "topicalization," "thematization," and "rhematization." Its discussion of spoken discourse includes tone group, pitch, and tonic syllable. Cohesion is also discussed, with examples from the same publications, focusing on lexical cohesion, i.e. repetition, synonymy, antonymy, derivation, inclusion, and collocation. The paper is dependent on notions from the Prague School of Linguistics (Functional Sentence Perspective), M. A. K. Halliday's functional grammar, and research in the analysis of written discourse. Thirty-nine references and an appendix containing the sample texts are attached. (SR)

**********
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
**********

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
Kim Brian Lovejoy
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☐ This document has been reproduced as received from the person or organization originating it
☐ Minor changes have been made to improve reproduction quality
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy

ED310383

# Discourse Analysis: Part I, Information Management and Cohesion*

Kim Brian Lovejoy
Department of English
Indiana University/Indianapolis

Donald M. Lance
Department of English
University of Missouri/Columbia

# Discourse Analysis: Part II, A Comparison of Published Writing in Three Disciplines

Kim Brian Lovejoy
Department of English
Indiana University/Indianalopis

Running Head (Part I): Discourse Analysis: Information Management and Cohesion
Running Head (Part II): A Comparison of Published Writing in Three Disciplines

---

*For purposes of "blind" reviews, we are using "first author" and "second author." In the published version appropriate last names should be entered.

CS211956



BEST COPY AVAILABLE

# Discourse Analysis: Part I, Information Management and Cohesion

Kim Brian Lovejoy
Department of English
Indiana University/Indianapolis

Donald M. Lance
Department of English
University of Missouri/Columbia

## Author Notes

A preliminary version of this article was presented at the Fifteenth International Systemic Congress at Michigan State University in August 1988. Preparation of the article for publication was supported by summer research grants from the President's Council on Arts and Science and from the School of Liberal Arts, Indiana University/Indianapolis.

Correspondence and requests for reprints should be sent to Kim Brian Lovejoy, Department of English, CA 502J, 425 University Boulevard, Indianapolis, IN 46202.

---

Running Head: Discourse Analysis: Information Management and Cohesion

Abstract

Part I of this two-art article proposes a model for the analysis of information management and cohesion in written discourse. Concepts of discourse analysis are defined, specifically information management, syntax, semantic reference, lexicon, cohesion, and intonation, with examples taken from publications in psychology, biology, and history. Key terms related to reference and information include information unit, theme, rheme, topic, comment, given, new, focus, markedness, topicalization, thematization, rhematization. Discussion of spoken discourse includes tone group, pitch, tonic syllable. Cohesion also is discussed, with examples from the same publications. The discussion of cohesion focuses on lexical cohesion, i.e., repetition, synonymy, antonymy, derivation, inclusion, collocation. The article is dependent on notions from the Prague School of Linguistics (Functional Sentence Perspective), M.A.K. Halliday's functional grammar, and research in the analysis of written discourse.

Discourse Analysis: Part I, Cohesion and Information Management

Only in fairly recent years has research in linguistics attempted to contribute to composition studies. In his bibliographical essay, Strong remarks that "what was intellectually provocative" in linguistic studies twenty years ago "was nevertheless remote from everyday reality" (1985, 68). Larson (1979) shares this view in his assessment of the relevance of linguistic studies to the complex processes of writing. Prior to the mid-sixties, Larson points out, linguistics offered guidance to students in the editing stage of their writing, but even here the influence of linguistics on students' learning to compose was seriously questioned by the research on grammar instruction reported in Richard Braddock's (et al.) Research in Written Composition (1963). Teachers of writing, consequently, saw little of value in what linguistics had to offer.

The turning point came with Hunt's (1964) notion of the T-unit, a research mechanism that legitimized descriptive analyses of written texts, making them more accurate and hence more reliable. Hunt's research in textual analysis led to the establishment of "descriptive norms of written syntax" (cited in Strong 1985, 68) and commanded the attention of composition specialists because he provided a method of analysis that accounted for what teachers knew about writing but could not substantiate. By "operationalizing descriptive (and qualitative) studies," writes Strong, Hunt "provided a vital link between linguistics and writing" (1985, 69). For teachers of composition, this link between linguistics and composition is best exemplified

by what is perhaps the most significant outgrowth of Hunt's research in syntactic structure: sentence combining. As an instructional method, it is now sanctioned by empirical studies for its relevance to the teaching of writing. Subsequent research by Young, Becker, and Pike (1970) on tagmemic rhetoric and, more recently, by the text linguists (Kintsch, 1974; Halliday and Hasan, 1976, 1985; Dyke, 1979; Beaugrande and Dressler, 1981; and others) has been largely responsible for the increased interests in coherence over the last decade. Such research clearly affirms the role of linguistic studies in composition research today.

The present study combines linguistics and composition studies by proposing a descriptive model for the analysis of spoken and written discourse. The research was predicated on the notion that the concepts of discourse analysis (especially theme/rheme and given/new) and cohesion would be useful in analyzing styles and genres of discourse. One of our original assumptions was that the patterns of use of lexically cohesive elements and the distribution of information would show interesting variations in written texts from diverse disciplines. To test this hypothesis, we chose passages from published scholarly articles in three disciplines in which teachers often participate in programs in writing across the curriculum and in reading in the content areas (counseling psychology, biology, and history[1]): these sample texts appear in Appendix 1. In choosing passages from diverse disciplines, we hoped that the results of our analyses would be useful to teachers and researchers in composition, adding to an expanding body of

knowledge about writing in the disciplines.

Part I of this two-part article focuses on the principal concepts and terms used in the classification and analysis of information units and of cohesive devices; Part II summarizes and discusses the findings of the analysis of data from the sample passages. Because the publications on information management have covered a diversity of aspects of text analysis and offer a wide range of terminology--often different terms for the same basic concept, we decided to follow Halliday's definitions as closely as possible, making adjustments when the data posed problems that could not be easily handled with existing terminological designations. Because cohesion has been treated much more thoroughly in previous studies, fewer terminological and analytical questions arose in this latter portion of the model.

## Information Management[2]

Studies of texts tend to treat reference as as if it is strongly related to syntax, if not part of it. Our assumption, however, is that the dispensing of information and the use of language per se are separate components of an act of communication; information management is the superordinate component. When speakers and writers have ideas to share with others, they dispense these ideas as chunks of information embedded into linguistic units appropriate for the social or personal function of the communicative discourse. Two components of language serve as the primary means of dispensing these "chunks"--lexicon and syntax. The ideas constitute the referents

of the lexical items chosen for the discourse, and the speaker/writer chooses words and syntactic structures to indicate interrelations among the referents. In our model we use the term "information management"[3] to refer to how the speaker or writer distributes lexical items and syntactic structures within clauses and sentences in the process of "communicating ideas," i.e., in managing information for particular communicative purposes.

The basic notions of information management were first popularized by Vilem Mathesius, one of the members of the "Prague School" of linguistics, who introduced the term "Functional Sentence Perspective" (FSP) in an article in 1939; the two components of FSP are "a) the arrangement of the content structure of the sentence; b) its connection with the context" (Beneš, 1968, p. 267 and note 1). Firbas (1974), a Czech scholar who has studied FSP in English, traces the origin of FSP to an 1844 monograph by Henry Weil, a French classical scholar. The monograph, translated into English in 1878 under the title The Order of Words in the Ancient Languages Compared with That of the Modern Languages, contains many of the ideas that inspired Mathesius and others in the Prague linguistic circle (1974, 11). According to Firbas, Weil attempted to show that "men think and express themselves in the same order whether they speak a modern language or use one of the ancient languages" (1974, 12). Firbas provides this synopsis of Weil's seminal work:

> A sentence contains a point of departure (an initial notion) and a goal of discourse. The point of departure is equally present to the speaker and to the

> hearer; it is the rallying point, the ground on which they meet. The goal of discourse presents the very information that is to be imparted to the hearer. ... the movement from the initial notion to the goal of discourse reveals the movement of the mind itself. Languages may use different syntactical constructions, but the order of ideas remains basically the same.
>
> (1974, 12)

Over the years, linguists have produced an abundance of terminology in attempting to account for certain aspects of the communicative function of the sentence. Certain aspects of FSP theory, for example, have been referred to as "point of departure" and "goal of discourse," "theme" and "rheme," "topic" and "comment," "given" and "new," "known" and "unknown," "old" and "new," "presupposition" and "focus," etc. (see, e.g., Firbas, 1966, 1974; Daneš, 1974; Vachek, 1974; Palkova and Palek, 1978). The terminology and point of view of our study derive primarily from Halliday (1967a), who cites the works of Danes and Firbas in his early publications on theme and text analysis; however, we have felt the need to make substantial modifications because our analysis of published expository discourse requires a different approach from that used in earlier studies, most of which focused on individual sentences or small sets of sentences from interactive conversation.

The study reported in this publication makes use of several features of information management and several aspects of lexical cohesion. The key terms of information management, explained

below, are theme/rheme, topic/comment, given/new, marked/unmarked, and information focus, with the most important being theme/rheme and given/new. In essence, a theme is a set of words in sentence-initial position that indicate "what the sentence is talking about," and a rheme contains information about its theme that the speaker/writer wants the listener/reader to consider at a particular point in the development of a text or conversation.

Studies conducted in the past have tended not to differentiate between syntax and information management; they have fused meaning/pragmatics (i.e., reference) and grammatical analysis. Generally speaking, others have conflated the pairs of information units that we have labeled as "theme/rheme" and "topic/comment"; both the Prague linguists and Halliday (1967a, 1985) use the former pair in their analyses, whereas the second pair is more popular among other linguists. In a recent comprehensive grammar of English, Quirk et al. (1985, pp. 1355-77) briefly review the varying uses of theme/rheme, topic/comment, given/new, etc. and settle on the pairs given/new and topic/focus in analyzing examples from spoken discourse (pp. 1360-64). Prince has studied the factors involved in the given/new distinctions in depth and points out that other researchers have used a variety of terms in discussing textual information, often employing the terms in such a way that "no two of them mean the same thing, and that, in some cases, the differences are quite large" (p. 225). She focuses on given and new and sets up a taxonomy of "givens" that is very convincing in

her analysis of brief interchanges of spoken discourse, but she makes some distinctions that we did not find useful in our analysis of published texts (e.g., anchored/unanchored and the distinction between inferrable and evoked). The categories that Prince describes would need to be included in thorough studies of a variety of discourse types and functions and in more comprehensive studies than ours.

In "ordinary" sentences containing subject-verb-object order with no special emphasis (i.e., the unmarked sentence), the theme, topic, and subject coincide, as do rheme, comment, and predicate; other orderings of constituents, however, are common in written discourse and must be given serious attention. Though the "usual" theme is a noun phrase, Halliday (1985, pp. 39-40) points out that other structures may be used as themes. Deviations from the "ordinary" ordering of constituents results in marked sentences. For example, when adverbial structures appear at the beginning of a sentence--as well as several other structures--the reference in the initial adverbial "sets the stage" for the information that follows, as in "Because the weather is so nice, let's have a picnic today." It is also often argued that the theme/topic of an unmarked sentence will contain given information and the rheme/comment will contain the new information, but Fries (in press) has demonstrated that themes may contain a variety of types of information. The example given here could serve as a conversation opener in which the interlocutors have not been speaking of either the weather or social activities; though the listeners may or may not have noticed the weather, the sentence is not likely to be spoken

unless there is a possibility for "picnic weather" to exist.

Fronting of sentence constituents to pre-subject position is usually called topicalization, though those who use the term theme refer to such fronting as thematization. Li and Thompson (1976) present evidence that the languages of the world can be typologically classified on the basis of whether the unmarked sentence has a topic or a subject in sentence-initial position. They conclude from their analyses of data from a dozen language families that "the topic is a discourse notion, whereas the subject is to a greater extent a sentence-internal notion" (p. 466); topic prominence is much more common in Oriental languages than in Indo-European and other language families (p. 460). In one form of Chinese topicalization, a topic noun phrase precedes the subject of the first in a sequence of predications containing interrelated ideas; the succeeding predications repeat neither the topic nor the subject until one or the other changes or unless there is a need for emphasis or clarification (see, e.g., Tiee, 1986, pp. 1-3, 328-36). Friedman (1976, pp. 142-146) has found in her research that the hearing-impaired use topic/comment divisions and topicalization in American Sign Language (ASL) in ways that are analogous to the ways in which these processes are used in spoken language; interestingly, when successive sentences have the same topic, ASL topic marking is similar to the Chinese use described by Tiee, pp. 1-3.

Chinese "topics," as well as "topics" as analyzed by Givon (1983) and others, are instances of what we term themes--i.e., in our model, in which information management is treated as separate

from and superordinate to syntax. Dik's analysis of constituent order in Portuguese (1981, pp. 170-171) supports our position that non-subject information placed at the beginning of a sentence and followed by a pause is closer to Halliday's defintion of theme than to the structures for which we use the term topic. The variation in terminology is so diverse, and each scholarly publication has so many valuable analyses, that we do not claim that in choosing certain terms and concepts over others we have "solved" the question of how the terms subject, topic, theme, etc. should be used in analyzing texts from all languages, or even all styles and text genres in American English; we have selected notions and terms from the literature that come closest to explaining the data at hand, and in these articles we present our case for other scholars to consider so that they may see how well our model works in analyses of other text types.

One of the principal features of our model is the separation of information management (a pragmatic function in interactive communication) and the grammar and lexicon of the sentence. Lexis serves as the interface between information and syntax and plays a significant role in cohesion, and it is relevant to information management when a lexical item in question plays a role in marking themic or rhemic information. As writers write or speakers interact with others, the producer of a text (or portion of an oral interchange) has information to share with an audience. In effective use of discourse techniques, the producer will organize relevant chunks of this information into themes and rhemes to keep the reader/auditor continually apprised of the topic of the discourse through careful selection of given or

known information in the themes of sentences and will distribute information new to the argument in the rhemes--or, if appropriate, in themes--in an attempt to accomplish the communicative goals that motivate the presentation of the information. Inexperienced users of discourse techniques often communicate less effectively than they intend because they inappropriately place in themic position information that is completely new to the reader, or because they do not maintain cohesive ties between successive portions of the argument. In either of these violations of effective information management, the producer has failed to "manage" themic and rhemic references so that the audience can easily follow the intended flow of given and new information. Witte and Faigley (1981) and Vande Kopple (1982, 1983) have found that students write better papers when they develop the ability to use new information more effectively in the topics discussed in their essays. Close attention to what the listener/reader already knows from prior experience, from the immediate context, or from previous references in the text is central to the process of information management.

In the analysis of written discourse, the chief unit to be considered is the sentence--i.e., in simplistic terms, what one conventionally writes with a capital letter at the beginning and terminal punctuation at the end, the "orthographic sentence." In our study, the clauses in compound sentences joined by coordinating conjunctions, conjunctive adverbs, or non-terminal punctuation are analyzed separately, along with any clausal or phrasal modifiers or complements. This approach echoes the

research of Hunt (1964, 1966), Christensen (1965), Larson (1967), and Winterowd (1970), who use the "T-unit" (terminable unit) as the basic independent structural unit in their analyses of information within texts. In the paragraphs that follow, we show how the key terms in our model apply to selected sentences in the texts that were used as the corpus of data for [the first author's] dissertation. Because this study is based on written texts, we will refer to "writer/reader," though in discussions of spoken or gestural communication obviously other terms would be more appropriate for the participants in an interchange of information.

In the management of information that goes into a text, the most useful concept for the writer is the theme-rheme division. The theme, as defined by Halliday (1967a, p. 201), is the "point of departure" for the presentation of information. The information that the writer wishes to impart about the theme constitutes the rheme; it is usually new to the argument at the point at which it is introduced into the text. The theme-rheme combination is "the basic form of organization of the clause as a message" (Halliday, 1985, p. 53; emphasis ours). The rhemic information may, of course, be old, known information that the writer needs to bring to the reader's attention again for present purposes, but at a given point in the text it may be new to the argument. For example, very old and very well known information could serve as "new" in a sentence such as, "During the confusing times of the Inquisition, it was of no small significance that Columbus proved beyond a doubt that the earth is round"; the informational function of the sentence is to point out the

significance (a focal point in the rheme) of the contemporaneity of the Inquisition and Columbus' voyage.

In his early works Halliday described the theme-rheme distinction as being marked by pitch contours. Though several researchers in psycholinguistics have conducted experiments that appear to refute Halliday's claim about pitch and themicity (summarized in Brown and Yule, 1983, pp. 159-69), we found the original definition useful--and true. The fact that a particular speaker may not always have an easily detected period of silence between the theme and rheme in a given sentence is a matter of style or of circumstances of delivery. In analyzing the three samples in our study, first we made tentative theme-rheme divisions on the basis of the information in each sentence in its context; then both of us independently read the sentences aloud several times, always keeping in mind the significance of each sentence in its context. We always found that in a careful reading there is a noticeable pitch drop at the end of the theme and that near the beginning of the rheme--often on the first word--there is an abrupt peak in pitch level, though this pitch peak is not necessarily on the tonic syllable of the tone group (defined below). There were only some minor variations in our renderings of the sentences in our samples; these occurred when there were words of low informational value at the break, e.g., noun clauses beginning with the subordinating conjunction <u>that</u> in a highly predictable subject-verb combination, as in "the author says that..." or "the study found that...."

Though published texts are not necessarily written to be

read aloud, neither are they written is such a manner that they could not be read aloud. Our approach assumes that if good performers read the texts aloud, with the meaning apparently intended by the authors, the theme-rheme divisions will fall where we claim, and there will be appropriate primary and secondary tone units in the longer sentences, with pitch peaks indicating not only unmarked focus but also such marked features as contrastive stress and certain types of informational focus. We recognize the potential for circularity in our procedure; but in the absence of the writer--who, unfortunately, is often less reliable than an attentive reader in expostulating on his own syntax--we had no other recourse.

A key component of the Hallidayan analysis of theme and rheme is intonation, explained in detail in Intonation and grammar in British English (1967b), with a somewhat elaborate taxonomy of tone contours associated with specific semantic and grammatical classifications of sentence types. In our analysis of published texts, we felt that we should use only a limited amount of his theory of tonality because we were not analyzing the same kinds of texts. In Halliday's analysis, tonality, the distribution of tone groups over stretches of speech, "can be regarded as the distribution of 'information units'" (1967b, p. 21). Each tone group consists of a pitch contour distributed over a group of words, with a major pitch peak on the tonic syllable of one of the words in the group; this pitch peak indicates the focus of the information unit. In unmarked utterances, the tonic syllable will occur in the last lexical item in the tone group, but in marked utterances attitudinal factors such as contrast,

doubt, reservation, hesitation, irritation, reproach, interrogation, etc. may alter either the tone contour or the location of the tonic, or both. For instance "Michael // is an acquaintance of mine" is an unmarked statement of fact, whereas "Michael is // an acquaintance of mine"[4] would be a marked sentence that emphasizes the truth of the assertion (see Quirk et al., 1985, p. 1415).

As indicated above, for our analysis we read each sentence of the corpus aloud to find the "pause" between the theme and the rheme. In our reading of short sentences like those in the preceding set, each sentence is a single tone group: in the unmarked sentence, there is slightly elevated pitch on Michael to indicate themicity, then a slight dip in pitch (theme-rheme division). The rheme begins with a low pitch that is sustained until the pitch peak on -quaint- in the last lexical word in the sentence, which is the point of informational focus for the sentence as a whole; then the coda of the tone group has lowered pitch, with a sudden drop at the end of the word mine. In the second sentence, the pitch begins low, then rises abruptly on is, rising higher than on the unmarked tonic syllable in the previous sentence, and then after is there is relatively low pitch except for a slight peak in mine. Each of these sentences is (hypothetically) a unit in a conversational interchange, an example of Halliday's observation that "one tone group is as it were one move in a speech act" (1967b, p. 30). There were many other features of tonality that we could have considered (e.g., tonicity, foot boundary, secondary tone (Halliday, 1967b, p. 53)),

but because of the scope of the study at hand we limited ourselves to this one intonational feature.

The first sentence from the psychology article is much more complex, intonationally and informationally, than these two sentences about Michael. When read in context, this long sentence also represents "one move" in the speech act that constitutes the entire published article[5]:

Research has demonstrated that // first impressions and stereotypes can influence social interactions in ways that lead to their behavioral confirmation--even to the extent of causing mistaken impressions to become real.

The sentence, when read aloud, is a complex intonational unit, with a theme-focal pitch peak on research and a pause after that. The highest pitch in the rheme is on even, but the rhemic stress is on real, indicating that "become real" is the focal information of the sentence as a whole; there are also several "ups and downs" in pitch contours on the sub-elements of the string of information units presented in the rheme. Halliday also points out (1967b, p. 22) that sentences may have more than one "major information point," signaled by more than one tonic syllable in the same tone-group sequence. His discussion is based on analyses of spoken discourse that does not have the complexity of embedding and multiplicity of information units found in the lengthy sentences of scholarly articles. We did not consider it necessary or desirable to designate the minor tone groups within the themes and rhemes of our texts in preparing the texts for statistical analysis, because we did not think that so much detail would

reveal anything significant, and we thought that excessive analysis would detract from our main goals; attention to these details could be of considerable interest in other types of analysis.

In his article on transitivity and theme (1967a, p. 200), Halliday states that the notion of 'topic' overlaps completely with 'theme' and 'given' and is therefore not included in his system. In our close analysis of formal published academic prose, however, we found three sets of distinctions to be useful and have coded the data accordingly. In the distinctions we are making, theme/rheme applies to information in an entire sentence, no matter how complex (with compound sentences divided into two (or more) units); topic/comment functions at the clausal level, applying to main clauses, subordinate clauses, and verbal phrases. Within a theme or a rheme there may be subordinate clauses, as we see in the sentence quoted in the preceding paragraph. The theme of the entire sentence is that research has demonstrated something. The pitch contour with which the sentence is read marks the sentence as containing two major units: the part before <u>first impressions</u> and the part beginning with this phrase. The reader will "pause" somewhere between <u>demonstrated</u> and <u>first</u>; that is, there will be a distinct drop in pitch at this point in the sentence. The pitch may fall to its lowest level either before or after <u>that</u>, but since this word serves only to introduce the noun clause and carries no content, the ambiguity over whether <u>that</u> is part of the theme or the rheme is informationally vacuous. The reader will note that in this

sentence the theme/rheme division does not fall between the subject and the predicate; such is often the case.

Within the rheme of this complex sentence we have a subject-predicate relationship in the noun clause ("first impressions and stereotypes" + "can influence . . ."), as well as a relative clause, a gerund phrase, and an infinitive phrase, each with a subject ("[these ways] lead to their behavioral confirmation"; "[first impressions and stereotypes] cause mistaken impressions to become real"; "[mistaken impressions] become real"). The subjects of the two subordinate clauses and the "understood" subjects of the two verbal phrases provide the informational topics about which the predicates make comments.

In the literature on discourse analysis, both theme and topic have been used to designate, in Halliday's words, "what we are talking about"--as has the term subject in traditional grammar. While "demonstrated" is the key element in the theme of the entire sentence above, someone did the demonstrating, and it was researchers who did so (with scientific prose relegating the human beings to obscurity via the agentless passive and the metonymous noun research). Thus, we needed to designate research as being what the main clause is "talking about," though the sentence as a whole is "talking about" the reference in demonstrated. Rather than using the term theme for clause-initial syntactic elements that have been embedded into themes and rhemes (i.e., structural and topical themes), as Halliday does in his Introduction to Functional Grammar (1985, pp. 53-56), we felt that the terms topic/comment were appropriate for "topical themes," particularly in noun clauses like "research has

demonstrated that NP+Verb...," a formulaic way of beginning an article in a scientific journal. Our use of both theme/rheme and topic/comment enabled us to make some interesting observations about clause structure in the three disciplines included in our study; see, for example, the discussion of syntactic complexity as demonstrated in Table 7 in Part II of this publication.

From a themic perspective, in the introductory review of literature on stereotypes and self-fulfilling prophecy, the authors focused on what was demonstrated rather than on the research projects themselves. We did not want to use the term subject as the term for the information-processing role of impressions and stereotypes, because that term refers to a grammatical category; we needed a term that reflects informational function but differentiates between theme and the syntactic function of the structure in question.

One can also see in the preceding example that the topic may be unspecified and informationally relevant, but of low relevance; i.e., the "understood" subjects/topics of the gerund and the infinitive can be recovered from the text and thus are "meaningful" to the reader. Such is not always the case, however, as in the second sentence in the third paragraph of the psychology article in which the identity of the referents of the grammatical subjects of the gerunds is of no informational relevance:

[someone] Describing a person as seeking psychological therapy, like [someone] labeling a person as being mentally ill, // implies that the person has psychological problems and is incapable of handling his or her own problems.

This sentence also has a complex theme. The first gerund phrase, with a pitch peak on therapy, is followed by a prepositional phrase referring to preceding discussion. The second phrase of the theme ("like labeling...") would be delivered orally with lowered pitch; its informational role is to remind the reader of the discussion in the preceding paragraph.

The opening sentence in the biology article presents an interesting challenge for the analysis of themicity:

Polonium-210 and lead-210, members of the natural uranium series, // are found in cigarette tobacco.

Brown and Yule (1985, p. 155) claim that texts must begin with new information; this example, however, demonstrates that the given/new constraints of the opening sentence of a text are more complex than their bold (i.e., simplistic) claim. The task faced by a writer struggling with an opening sentence is more realistically described by Beneš in his discussion of Functional Sentence Perspective: "the introductory sentence either linguistically actualizes the situation or somehow replaces (simulates) a context...: thus the theme is presented as something given" (1968, p. 269). In their opening sentence, the authors of the biology passage announce to their readers that they are going to talk about certain isotopes of two trace elements, which they assume the readers will recognize as being members of the natural uranium series. In reading the sentence aloud, we found that we did not have a "dramatic pause" until the end of the appositive phrase, an indication that the complete subject, including the

appositive, constitutes the theme. Readers of the journal in which the article appears, Toxicology and Applied Pharmacology, may not be able to predict that a particular article will begin by referring to polonium and lead, but they would not be surprised to find these nouns in the subject of the opening sentence of an article, particularly when the form of reference suggests radioactivity. The writers also assume that the readers probably know that these particular isotopes are members of the natural uranium series but feel that it would be wise to remind them of this fact. Thus, when one reads the sentence, one places themic pitch on polonium and lead, then drops the pitch for the appositive and pauses after series; the remaining words then are read with a slight pitch peak on found and the tonic pitch of the sentence on cigarette. Non-scientists may, on the first attempt, read the sentence differently because they do not share the same background information as the writers, but this sentence will "work" as an opening sentence only if read with the complete subject considered as presupposable within the reader's prior knowledge. Evidence of these presuppositions is given in the second sentence when the authors use the standard scientific abbreviation "$^{210}$Po," which a nonspecialist would not necessarily be expected to interpret immediately. The focus of attention for the opening sentence is not on polonium and lead but on "cigarette tobacco"; with the theme and focus established as they are, the authors may then relate what happens to these radioactive trace elements after subjects (in their experiment, rats) inhale cigarette smoke.

In dispensing information, a writer or speaker must keep in mind what is "given" and what is "new" to the audience. In unmarked sentences the theme will contain information that is "given," i.e., known by the reader or assumed to be known. As indicated above, each article begins with references to information that the writer assumes the reader would know. Because nothing precedes the first sentence, the "givenness" derives not from prior mention but from what we term "presupposable information." Such would not necessarily be the case, however, when the title of the article provides the informational context for the first sentence; in this situation, the themic information would be "textually given." The tendency for the writer to use to advantage assumptions about what is given, or already known to the reader, contributes to what the Praguian linguists call "communicative dynamism" in Functional Sentence Perspective (Firbas, 1966, pp. 270-276; Quirk et al., 1985, pp. 1356, 1394, 1431).

A third type of givenness is contextual. The latter two types of given information--textual and contextual--will be illustrated in the following paragraphs. The psychology article opens with the sentence cited above in which the writer tells the reader that research has demonstrated that first impressions and stereotypes can influence people's behavior. The second sentence expands the discussion by referring to a specific study that yielded such results:

In one study, for example, // Snyder, Tanke, and Berscheid investigated the process of behavioral confirmation of the

stereotype associated with physical attractiveness.

The head word in the theme of this sentence is study, a reference that can be presumed to be one of the studies in the research mentioned in the opening sentence. Though at this point the study is not "known" information, it is a pragmatically given in view of the opening sentence of the article: research consists of studies. At this point, the authors wanted to relate the results of the study done by Snyder and his colleagues, but because the identity of the researchers is not "given" and is not presupposable, the authors need some means of setting up a "given" reference; thus they choose a word directly associated with research as the main word in the theme of the opening sentence. After the authors have established the fact that they are reporting on research by several different psychologists, they may operate with new names as if they are given, following Benes's observation that a writer may present a theme "as something given" under certain circumstances. In the psychology passage, in the third and fourth paragraphs, when the authors want to introduce new studies, they do not need to set up a convenient "given" theme but may treat the new names as if they are given: "Phillips, for example, found that...," "Piner and Kahle demonstrated that...."

The majority of the topics in the sample passages of our corpus have textually given information. Representative examples are the last two (sixth and seventh) sentences of the first paragraph of the psychology article:

Snyder, et al. concluded that // male perceivers used different

styles of interaction for the two groups of targets. These behaviors, in turn, // guided and constricted the behavioral options of female targets in ways that led them to conform to the men's initial impressions.

In these two sentences Snyder and behaviors are, respectively, repetitions of references in the fourth and sixth sentences of the paragraph and thus are clearly textually "given."

Generally speaking, as Clark and Haviland (1977) explain in their description of the "given-new contract" in conversations, rhemes contain information that is new to the argument. The first paragraph of the psychology passage has only new information in the rhemes, not surprising in view of the need to establish a framework of reference for succeeding paragraphs. The second paragraph introduces a second study on first impressions and stereotypes, opening with a textually given theme so as to tell the reader that the two paragraphs are directly related:

The reasoning, outlined by Snyder et al., // is similar in many ways to the processes proposed by Becker and Scheff relating to the labeling approach to social deviance.

Now that Becker's and Scheff's studies on labeling theory have been introduced, the authors can begin the next sentence with that term:

Labeling theory // suggests that once a person is labeled as mentally ill, preexisting stereotypes are activated in other people.

From this point in the text, the authors need to embed given information in the rheme to indicate how the passage on Becker's and Scheff's studies are related to what has been said about the study by Snyder et al., as in the third and fourth sentences (given information underlined):

With respect to mental illness, // the public generally perceives mental patients and ex-mental patients as threatening and socially undesirable.
According to labeling theory, // based on these perceptions, people systematically alter their expectations, vocabulary, and response cues when they interact with mental patients or ex-mental patients.

In the third sentence "the public" refers directly to people in the preceding sentence, and "perceptions" refers to perceives in the preceding sentence and echoes the perceptions discussed in the Snyder et al. study.

As the authors begin the third paragraph of the psychology article, they employ very little given information, perhaps to signal that they are shifting from research of one type (on mental illness) to another type (on psychological therapy administered in situations less severe than "mental illness"). The paragraph begins with a sentence that has "dummy topics," i.e., with initial expletives used in two different clauses so that the "subjects", new information, may be placed later in the sentence:

Although there has been little systematic investigation, // it appears that seeking psychological therapy may be associated with

being mentally ill.

The preceding discussion covers the principal concepts of information management that we used in the analysis of our corpus of data. Before continuing with the final set of terms (types of informationally marked sentences), we feel that at this point the reader would benefit from a reiteration of the definitions of the main terms that have appeared in the preceding pages.

INFORMATION UNIT: A group of words that function together as a unit within the sequential flow of information in a text. Each information unit is spoken with a single intonation contour and has one word (or more than one, in complex units) that is stressed, with pitch being the prominent feature of informational stress.

UNMARKED INFORMATION UNIT: A group of words arranged linearly in the "ordinary" sequential order common in a language: in English, clause = subject - verb - object - adverbial; noun phrase = determiner - adjective - prepositional phrase / relative clause. The unmarked structure is spoken with a single tone contour, with focal stress on the last lexical item; it may also contain secondary information groups within it.

MARKED INFORMATION UNIT: A group of words that deviates from the unmarked order of constituents or that has focal stress and pitch different from the "ordinary" intonation of an unmarked unit containing the same words. Three major types of markedness are Thematization, Rhematization, and Pseudo-

Thematization (see below).

THEME: "What the sentence in talking about"; "the point of departure" for the discussion in the rheme; the structure (or structures) at the beginning of a sentence. The theme is read or spoken with a single subordinate tone contour within the intonation of the entire sentence; the theme begins with relatively low pitch, preparatory to a pitch peak on the word (or words) that mark the informational focus of the theme, and then the pitch is lowered perceptibly after the theme-focal word and remains low until the sharp pitch drop that marks the theme-rheme boundary.

RHEME: The information about the theme that the writer wants the reader to add to the argument at a particular point in the progress of the argument in the text. The rheme is read or spoken with a continuous tone contour but may contain more than one subordinate information unit within it, the latter marked by slight rises and drops in pitch.

TOPIC: The initial constituent in a predication with unmarked word order (main clause, subordinate clause, verbal phrase); it serves as a syntactic analogue to the (informational) theme. Though not always spoken with themic intonation, the topic provides the "point of departure" for the information that follows in the predication.

COMMENT: The constituents of a clause that give information about the topic. Like rhemes, comments may contain more than one information unit, with intonation applied accordingly.

GIVEN: A reference that the speaker/writer may assume to be known or knowable.

PRESUPPOSABLE GIVEN INFORMATION: Information that may be assumed to be known by any member of the discourse community that serves as the audience for a spoken or written text.

CONTEXTUALLY GIVEN INFORMATION: Information directly or indirectly related to the subject matter of a written or oral text.

TEXTUALLY GIVEN INFORMATION: Information that has already been mentioned in the text; the earlier mention may or may not have been previously known by the audience. The repetition may occur in a synonym, a paraphrase, or a related word or expression.

NEW: Information that the speaker/writer provides in a text for purposes of furthering the argument. The information may already be known to the audience but not yet specifically associated with the argument.

MARKEDNESS: Sentences are informationally "marked" if constituents are not placed in their expected positions or if the sentence focus does not occur on the last lexical item. Three types of marking that occur in our corpus are thematizations, rhematizations, and pseudo-thematizations. Whenever any of these processes is used, the information in question is placed where it is in order for the text to serve a particular communicative purpose.

THEMATIZATION entails placing in initial position a structure other than the nominal group that serves as the subject of the sentence; such structures are instances of thematizations only if they are pronounced with themic intonation and serve the

informational function of "setting the scene" for the rhemic information in the latter part of the sentence. Fourteen instances of thematization occurred in our corpus, in the following syntactic structures: seven prepositional phrases[6], four adverbial clauses, one adjective phrase as an absolute construction, one gerund phrase, one participial phrase.

In our corpus, all except one of the initial prepositional phrases has given information that serves as a link to preceding discussion: "In one study...," "With respect to mental illness...," "In about one of three smokers...." In each of these phrases the object of the preposition also forms a cohesive lexical tie by repeating at least one previous reference. In the only thematized prepositional phrase with new information, the object (tests) is not cohesively tied to a preceding reference, but post-modifiers in the phrase have nouns that constitute cohesive ties to preceding sentences:

By the usual tests of the freedom of the practitioners to govern entry and exit from the field..., // journalists are not as autonomous as, for example, physicians and attorneys.

Though the words practitioners and field are members of chains of cohesive ties in references to journalists and journalism, this opening phrase about tests of freedom introduces a new idea in the development of the argument of the essay.

The four sentences with intitial thematized adverbial clauses are similar to the prepositional phrase with new information in that they give a situational setting for the following rheme. Two of the themic adverbial clauses specify conditions under which

the rhemic information applies ("Although there has been..."; "If they sought..."), one gives a temporal-conditional setting ("When a cigarette is smoked,..."), and one follows a prepositional phrase to indicate a combination of location and condition for the following rheme:

In the Soviet Union, where journalistic practice must reflect party-mindedness, ideological orthodoxy, and political loyalty, // how thoroughly are the working rules of journalism pervaded by general principles laid down by party doctrine?

RHEMATIZATION entails placing in a rheme a structure that also could be used as the subject of the sentence and, in appropriate circumstances, the theme of the sentence. Such structures are rather complex and consequently do not occur often. Of the 63 sentences in our corpus there were only three instances of rhematization: two with extraposition of noun phrases (in the psychology article) and one with extraposition of an infinitive phrase (in the biology article). One effect of rhematization is that it places most of the referential elements in focal positions. In extraposition, the subject of the main clause (and topic of the theme) is the informationally empty expletive <u>it</u>; thus the remainder of the theme is in the comment, and the structure for which <u>it</u> has been substituted is the rheme of the entire sentence, as in these two successive sentences from the psychology article:

It was hypothesized // that the perceiver would form more negative impressions of the partner when the target was believed

to be seeking psychological aid than when no mention was made about the target seeking therapy.
Furthermore, it was predicted that // the perceiver would interact with the target in ways that would lead to changes in the target's behavior that would confirm the perceiver's initial negative impressions.

These two sentences appear just before the concluding sentence of the introduction. They lead into the conclusion very effectively by placing themic focus on hypothesized and predicted and placing in final focal position the phrases seeking therapy and initial negative impressions.

PSEUDO-THEMATIZATION entails placement of focal pitch on a modifier of the subject in what otherwise might be an unmarked sentence, or special stress (e.g., contrast) on the head word of the subject. The prenominal structures that receive focal pitch are quantifiers, pre-determiners (another type of quantifier), and modifiers. Other structures, e.g., a word in a post-nominal modifier, could receive special stress in pseudo-thematization. The following are typical examples; stressed words are underlined:

Ninety percent of the lung cancer...
Most of the $^{210}Po$ in cigarette smoke...
A few measurements...
The present experiment...
The work reported here...
The Western ideal...

The soluble $^{210}Pb$...

The pseudo-thematized modifiers in the last four examples above receive the focal stress because they provide contrastive information. The one sentence with stress on a head noun (transfer) is marked because the authors have employed a compound subject but need to mark one of the head words as the focal element. The next sentence repeats the word transfer, and the whole article is about what happens as radioactivity is transferred from the smoke to different parts of the respiratory system. The following are the second and third sentences of the biology passage:

The transfer of $^{210}Po$ to the smoke and its presence in samples of bronchial epithelium from cigarette smokers were documented by Radford and Hunt.
When a cigarette is smoked, about 10% of the 210-Po is transferred to the mainstream smoke.

The discussion throughout this section on information management has repeatedly interlinked lexical cohesion and both themic and rhemic structures and given and new information. Now we turn to a discussion of the principal terms of cohesion and how we applied them to our sample texts.

Cohesion

The most comprehensive and integrative theoretical study of textual cohesion is M.A.K. Halliday and Ruqaiya Hasan's Cohesion in English (1976). Cohesion in Halliday and Hasan's sense of the

term refers to the lexical and grammatical means by which meaning relationships connect the content and structure of a text from sentence to sentence and from paragraph to paragraph within a discourse. These relationships function as bonding mechanisms in contiguous and non-contiguous sentences and contribute to what Halliday and Hasan have termed "texture" in writing (1976, p. 2). When an element in a text depends on another element in a preceding sentence and is interpretable only through the presence of an overtly expressed linguistic connection between the two elements, an instance of textual cohesion occurs--what Halliday and Hasan refer to as a "cohesive tie" (p. 3). In the sentence "They told her about it," for example, three elements (They, her, it) have meaning only if they are understood in relation to referents in a context. Taken out of context, as it is here, the sentence is referentially vacuous. If this sentence follows "The Smiths asked their accountant what Mary should do about a certain tax deduction and then phoned her immediately," one is able to assign probable meaning to the three pronouns. Though the use of anaphoric pronouns (grammatical forms substituted for preceding referents) is a clear example of cohesion, according to Halliday and Hasan, in extended discourse, it is the "non-structural text-forming relations" rather than grammar itself (p. 7) that contribute most to the cohesiveness of a text. Thus cohesion is conceived as a semantic or communicative phenomenon rather than one primarily of grammatical structure; "grammar" in the traditional sense is of only limited use in the study of intersentential relations.

Halliday and Hasan classify and sub-classify the types of

textual connections and devise a method of coding cohesive ties in discourse. Their coding scheme laid the groundwork that composition specialists have used in numerous subsequent studies. For the most part, these studies have focused on determining the relationship between cohesion and the quality of student writing. After almost a decade, however, this type of research has not provided clear and revealing results delineating a direct relationship between the number and types of cohesive ties and writing quality. Some report a direct correlation between the frequency of certain types of cohesive ties and writing quality, while others find no relationship whatsoever (compare, e.g., Witte and Faigley (1981) and Tierney and Mosenthal (1983)). One explanation for these contradictory findings may be the variations in the methodology used to interpret cohesive relations.

Researchers discovered early on that Halliday and Hasan's cohesion model was not easily adapted to the analysis of expository texts. Because Halliday and Hasan analyzed samples of conversation and literature, composition researchers who have used their model have been forced to collapse certain categories and to make adjustments in the method of coding ties. Stotsky (1983) was the first to make substantive changes in the system of categories proposed by Halliday and Hasan in order to account more effectively for the types of lexical cohesion in expository writing; she also enumerated the specific guidelines that she used to code lexical ties. In our analysis, we have followed Stotsky's guidelines for coding ties, but we also make further

refinements in order to reveal the richness of texture in published writing.

Since cohesion in writing functions to make explicit the information in a text, we felt that it is essential to combine analyses of both information management and cohesion to understand how texts "work." We chose to examine published writing in different disciplines in order to get a preliminary view of some of the major diferences in the ways in which theme/rheme, topic/comment, given/new, and lexical cohesion are used in texts of similar types in different academic disciplines, namely publications on the findings of research typical of each discipline. Whereas cohesion devices serve to connect elements in successive sentences in a discourse, information management accounts for how references are ordered within each sentence as the writer presents the argument central to the text of the discourse. Cohesion and information management are both dependent on reference and referents, whether related to information that has previously been stated in a text or to knowledge that the writer assumes is already known to the reader. Because the information varies considerably from one discipline to another--from specific, precise names for elements in the physical sciences to subjective notions in the humanities and social sciences that may be expressed in a number of ways--studies of how accomplished writers in academic disciplines use references and referents informationally and cohesively should yield valuable data for research and teaching in the fields of composition and reading.

The analytical model used in the study (i.e., the "layout" of data) is adapted from Halliday and Hasan's cohesion model in

Cohesion in English (1976, pp. 329-55), with adaptations by Sandra Stotsky (1983, 1986). We rely mainly on Stotsky's 1983 revision of Halliday and Hasan's lexical model, because her reformulation of their format provides a clearer description of specific textual features that occur frequently in expository prose. The model is divided into seven information types: (1) the sentence number in sequential order, (2) the total number of ties in the sentence, (3) the cohesive item, (4) the type of tie, (5) the referential antecedent in a preceding sentence or in the context, (6) the sentence number containing the antecedent, and (7) in the case of intratextual ties, the number of words entering into the tie. This last type was added by Stotsky (1983). Our coding of information units and cohesive ties are discussed and illustrated in Part II, with statistical tables displaying differences found in analyses of data from the three sample texts.

Halliday and Hasan (pp. 333-338) classify the various types of cohesive relations into five major categories (subcategories given in parentheses):

1. Reference (pronominals, demonstratives and definite articles, comparatives)
2. Substitution (nominal, verbal, clausal)
3. Ellipsis (nominal, verbal, clausal)
4. Conjunction (additive, adversative, causal, temporal, 'continuative', intonation)
5. Lexical (same item, synonym, superordinate, 'general' item, collocation)

For a brief but excellent overview of Halliday and Hasan's taxonomy of cohesive ties, see Witte and Faigley (1981).

In the analysis of our corpus, we found cohesive relations involving reference, conjunction, and lexical cohesion and their respective subsets in all the sample texts; however, as expected in expository texts, no cohesive relations involving substitution were found, and only the type of ellipsis that Halliday and Hasan refer to as nominal ellipsis was used cohesively in the sample passages. Because of our interest in information management, lexical cohesion was of somewhat more interest to us than the other categories. In our data, perhaps because of their referential nature, lexical ties constitute the majority of cohesive ties, as was the case in the studies of expository prose by Witte and Faigley (1981) and Stotsky (1986).

Referential ties consist of function words that refer to another element in the text, usually in preceding text but occasionally in references that follow, and often to the situation or discourse context. Halliday and Hasan divide the class into two subclasses: exophora, items that refer to the situational context; endophora, items that refer to the textual environment (p. 33). Endophoric ties may be either anaphoric, referring to preceding text (p. 14), or cataphoric, marking a reference that is clarified by a following modifier in the same sentence (p. 17). Only anaphoric references function in intersentential cohesion. In the sequence "Cigarettes are bad for you. They cause cancer," the presumed reference of they is "cigarettes," but if this sequence is preceded by "Cigarette

tobacco contains carcinogenic tars," the specific reference of they is not so easily determined: Do the cigarettes or the tars cause cancer? Halliday and Hasan (p. 314) define reference as a presupposition of co-reference; in ambiguous sequences like the present hypothetical example, the listener has no problem with "meaning" unless there is a need for precision of reference.

Substitution uses various types of words, such as one or some for nouns, do/did for verbs, and it/that for clauses or combined structures. A hypothetical example, not in our texts, would be "East Bloc workers from different regions get what they want in different ways. Some never get it, Ukranians use innuendo to get it, and Poles do so by going on strike." Because substitutions are more common in less formal texts, there were no instances in our samples.

Elliptical ties are deletions that signal by their absence that the reader already knows the reference. Because of the nature of the subject matter and the informational density of the texts in our corpus, we found instances of only one of the subtypes of this category of cohesion: nominal ellipsis, all but one of them in the biology passage. The first two sentences of the biology article offer a good illustrative example:

Polonium-210 and lead-210, members of the natural uranium series, are found in cigarette tobacco. The transfer of $^{210}Po$ to the smoke and its presence in samples of bronchial epithelium from cigarette smokers were documented by Radford and Hunt.

The reference of smoke in the subject of the second sentence, and

the definiteness of the noun phrase, can be understood only if the phrase means "the smoke of cigarette tobacco," a direct reference to the last two words in the preceding sentence.

Conjunctive ties indicate some sort of logical relationship between the content of one sentence and a preceding reference. Examples from our data are as follows:

A self-fulfilling prophecy thus occurred.

These behaviors, in turn, guided....

Phillips, for example, found that....

Furthermore, it was predicted that....

At the same time the term professionalism....

Lexical ties are repetitions of the same or related semantic references from one sentence to another. Halliday and Hasan coded only individual lexical items in their analysis, but Stotsky (1983, p. 437) found that lexical cohesion could be used effectively in the comparison of the language of good and bad student writing only if in counting ties one includes individual words, phrases, and even entire clauses. Stotsky also found that in analyzing expository prose she could make more usful distinctions by using the term inclusion rather than Halliday and Hasan's term superordinate, because the direction of cohesion may be from superordinate category to subordinate or vice versa (pp. 434-437). She also found that she needed to make some adjustments in the way she coded the types of references in the last two subtypes of Halliday and Hasan's listing of lexical cohesion.

The categories that we used in our analysis are basically those in Stotsky (1983), as follows, with definitions and

examples below:

Repetition

Derivation

Synonymy or near-synonymy

Opposition or contrast

Inclusion

Collocation

In discussing repetition, Halliday and Hasan (1976, p. 283) state that "it is not necessary for two lexical occurrences to have the same referent...in order for them to be cohesive." The psychology article provides a fortuitous test of this claim. Each of the four paragraphs in the introduction discusses a different type of experiment, but some identical words are used throughout: first impression, stereotype, behavioral confirmation, mistaken impression, target. The differences are not just a matter of identity of a person in an experiment about whom a "first impression" is formed: in the first paragraph the "targets" are individuals to whom physical attractiveness (or lack of it) is attributed, in the second the "targets" are previously institutionalized mental patients, in the third they are individuals who have sought psychological assistance from a clergyman or a psychiatrist, and in the fourth they are students whom the experimenters do or do not specify as having visited a university counseling center. Linking together all instances of the "same word" or one related to it in a single chain through all four paragraphs would do a disservice to the complexity of the texture of the passage. We preferred to separate the chains

referring to the separate types of individuals but to code as long chains those terms that were independent of references to specific target types. The difference between the cohesive function of (a) "first impressions" in reference to attractive targets and of (b) "first impressions" in the article as a whole has not been discussed in the literature to our knowledge, but this distinction might be labeled as (a) intratextual / (b) supratextual or (a) microtextual / (b) macrotextual cohesion. Lexical cohesion ties the passage together on two levels; however, because the references in the sets of chains are not identical, the coding of cohesive ties should not confound the levels.

Halliday and Hasan (p. 291) counted as the "same" lexical item derivationally related terms (nominal, nominalize) but not cognates (tooth, dental), but Stotsky (p. 432) felt that it would be valuable to set up a different category for derivations. This distinction was particularly useful in our model in analyzing the history passage, which contained derivationally related sets of words such as profession, professional, professionalism, professionalization. Pairs of words like tooth/dental would be coded as instances of cohesion or collocation (discussed below), depending on specific references.

Synonymy is similar to the preceding category, but should be treated separately not only because different lexical items are used but because the functions of the two types of repetition are different. Synonyms are connotatively different from each other when used in even slightly different contexts. For instance, in the history passage, in the second paragraph, news and information

refer hypothetically to the same people and events, but the two words could not be interchanged and have the same effect on the reader: "...the sociology of news...," "...the 'headline' approach to information...." Similarly, in the same paragraph, the "...general principles..." are the same as the "...day-to-day norms...," but the connotations of the terms keep them from being interchangeable in these two sentences.

The category "opposition and contrast," another of Stotsky's adjustments to the Hallidayan set of terms (p. 436), enables the researcher to make another useful distinction. Oppositions such as soluble and insoluble in the last paragraph of the biology passage are obviously necessary in scientific descriptions, as are attractive and unattractive in the psychology passage, but how does one refer to the opposite of journalist? In the history passage, the author chose the term outsiders to refer to non-journalists who would like to be reporters but have not established that they meet the "standards of the profession."

Inclusion entails the use of superordinate and subordinate members of ordered or unordered sets. In the psychology passage, this category was useful not only for obvious pairs like research and study but also for references such as actual dyadic social interaction and social interaction. In the biology sample, it was useful for references like airway surface, bronchial surface, mucosal surface, and bronchial epithelium.

Halliday and Hasan used the category collocation to encompass a much broader range of related terms than Stotsky does. She separated the cohesive categories into two major groups (p. 441): semantically related words (the five categories described above)

and collocationally related words. The latter category includes words that are cohesive in a text because of frequent co-occurrence in similar contexts; the cohesive power of collocations derives from the following four factors: frequency of occurrence in the language as individual words, frequency of co-occurrence in texts in general, physical proximity in the text, extent of reader's reading experience with the words. Two extended sets of collocational references were used in the biology passage: lung, broncial, epithelium, basal cell, parynchema and several words related to tobacco, smoke, and cigarette.

The examples selected for our definitions give an indication of the richness of the informational and cohesive texture of the three articles selected for our corpus. Accomplished writers and editors must master the techniques of managing information and controlling cohesive ties if they want their publications to be easily understood on first reading. As Stotsky points out in her 1983 article, however, the sophistication of the prior knowledge of readers will vary widely from one discourse community to another.

## Notes

[1]We gratefully acknowledge the assistance of Professors Hope Hills (psychology), James Carrell (biology), and Charles Timberlake (history), all of the University of Missouri/Columbia, who helped in selecting appropriate journal articles for the present study.

[2]Portions of this discussion on information management are heavily dependent on Lance (1989), an unpublished manuscript currently being revised.

[3]In many publications, the term "information distribution" is used where we use "information management" (e.g., Halliday, 1967a, p. 200); for connotational reasons, we prefer the latter term as the name of the process used by writers and speakers, though we will also refer to information being "distributed" within a text when we are focusing on the result or effect of the selection process used by the composer of the text. Cooper (1983) also uses the term "information management" in his review of linguistic approaches to the study of writing. The verbs distribute and manage are both process verbs with 'affected' participants as complements; however, the complement of manage has only the participant role of 'affected' participant in the predication (i.e., direct object), whereas the complement of distribute may fill either the single role of 'affected' or the compound role of 'affected-range.' Nominals and verbal phrases derived from these verbs carry the same role connotations as the verb + complement structures. (See Halliday, 1985, pp. 131-137,

for a discussion of participant functions.) The use of a 'process verb + affected participant' ("management of information") rather than a 'process verb + affected-range participant' ("distribution of information") focuses on the productive nature of language use in a communication event such as the writing of a technical article or a discussion among friends and consequently is a better term for the process of composing texts.

[4]In our examples in this essay, a double slash marks the division between the theme and the rheme.

[5]By deleting the bibliographical references in the examples within the text of this article we have in effect ommitted some of the information that professional readers use when reading for full information. Such references have a communicative function because they provide a co-text for readers who are or may be familiar with the authors cited and/or their research, and the dates allow readers to provide a chronological framework for the research cited. Information footnotes are particularly important to the reader's overall understanding of the author's meaning, as is the case with footnote 4 in the history passage.

[6]In [the first author's] dissertation, when the prepositional phrase contained given information, we considered the sentence-initial placement to be "topicalization (6 instances); however, since that time, we have decided that in view of their intonation and their informational function these structures should be classified as thematizations and subclassified as given- and new-thematizations. Thus, in its present form, our model has no structure that we would call "topicalization."

In a dissertation study in progress at the University of Missouri, using the same model (with some minor expansions) to analyze a conversation, a semi-formal lecture, and written texts by the same individual, Gudrun Boettcher Sherman has found that her preliminary data support our current position that the fronted prepositional phrases in the present study are instances of thematization. She points out (personal communication with the second author) that the thematization of given information serves as an informational transition ("Peter was in England from 1983 to 1986. During these three years, he attended Cambridge.") and that only one word in the sentence receives informational focus (Cambridge), marked by a pitch peak. When new information is fronted, there are two locations of informational focus--e.g., the following sentence spoken as a continuation of the previous two in the same discourse, "For two years before 1983, he studied at the University of Heidelberg." Thus, the fronting of new information results in a form of emphasis.

## References


Beaugrande, R., & Dressler, W.U. (1981). Introduction to text linguistics. New York: Longman.

Beneš, E. (1968). On two aspects of functional sentence perspective. In A.V. Isacenko (Ed.), Travaux linguistiques de Prague, 3 (pp. 267-275). University, AL: University of Alabama Press.

Braddock, R., Lloyd-Jones, R., & Schoer, L. (Eds.) (1963). Research in written composition. Urbana, IL: NCTE.

Brown, G., & Yule, G. (1983). Discourse analysis. Cambridge University Press.

Christensen, F. (1965). A generative rhetoric of the paragraph. College English, 16, 144-56.

Clark, H.H, & S.E. Haviland. (1977). Comprehension and the given-new contract. In R.O. Freedle (Ed.), Discourse production and comprehension (pp. 1-40). Norwood, NJ: Ablex.

Cooper, C.R. (1983). Procedures for describing written texts. In P. Mosenthal et al. (Eds.), Research on writing: Principles and methods (pp. 287-313). London: Longman.

Daneš, F. (1974). Functional sentence perspective and the organization of the text. In F. Danes (Ed.), Papers on functional sentence perspective (pp. 106-128). The Hague: Mouton.

Dijk, T.V. (1979). Macro-structures. Hillsdale, NJ: Erlbaum.

Dik, S.C. (1981). The interaction of subject and topic in Portuguese. In A.M. Bolkestein et al. (Eds.), Predication and expression in functional grammar (pp. 165-183). New York: Academic Press.

Firbas, J. (1966). On defining theme in functional sentence perspective. Travaux linguistiques de Prague, 1 (pp. 267-80). University AL: University of Alabama Press.

Firbas, J. (1974). Some aspects of the Czechoslovak approach to functional sentence perspective. In F. Danes (Ed.), Papers on functional sentence perspective (pp. 11-37). The Hague: Mouton.

Fries, P.H. (in press). Patterns of information in initial position in English. In P.H. Fries and M. Gregory (Eds.), Discourse in society: Functional perspectives. Norwood, NJ: Ablex.

Givon, T. (1983). Topic continuity in discourse. Amsterdam: John Benjamins.

Halliday, M.A.K. (1967a). Notes on transitivity and theme in English, Part II. Journal of linguistics, 3, 199-244.

Halliday, M.A.K. (1967b). Intonation and grammar in British English. The Hague: Mouton.

Halliday, M.A.K. (1985). An introduction to functional grammar. London: Edward Arnold.

Hunt, K.W. (1964). Differences in grammatical structures written at three grade levels. Tallahassee: Florida State University.

Hunt, K.W. (1966). Recent measures in syntactic development. Elementary English, 43, 732-39.

Kintsch, W. (1974). The representation of meaning in memory. Hillsdale, NJ: Erlbaum.

[second author]. (1989). Discourse: The managing of information. Unpublished manuscript.

Larson, R.L. (1967). Sentences in action: A technique for analyzing paragraphs. College English, 18, 16-22.

Larson, R.L. (1979). Language studies and composing processes. In D. McQuade (Ed.), Linguistics, stylistics, and the teaching of composition (pp. 182-90). Ohio: University of Akron.

Li, C.N., & S.A. Thompson. (1976). Subject and topic: A new typology of language. In C.N. Li (Ed.), Subject and topic (pp. 457-489). New York: Academic Press.

[first author]. (1987). A model for the analysis of cohesion and information management in published writing in three disciplines. Diss. University of Missouri/Columbia. Ann Arbor: UMI, 1988.

Palkova, Z., & Palek, B. (1978). Functional sentence perspective and textlinguistics. In W.U. Dressler (Ed.), Current trends in textlinguistics (pp. 212-227). New York: de Gruyter.

Prince, E.F. (1981). Toward a taxonomy of given-new information. In P. Cole (Ed.), Radical pragmatics (pp. 223-55). New York: Academic Press.

Quirk, R., Greenbaum, S., Leech, G. & Svartvik, J. (1985). A comprehensive grammar of the English language. London: Longman.

Stotsky, S. (1983). Types of lexical cohesion: Implications for developing a vocabulary of academic discourse. College composition and communication, 34, 430-445.

Stotsky, S. (1986). On learning to write about ideas. College communication and composition, 37, 276-293.

Strong, W. (1985). Linguistics and writing. In B.W. McClelland and T.R. Donovan (Eds.), Perspectives on research and scholarship in composition (pp. 68-86). New York: MLA.

Tiee, H.H. (1986). A reference grammar of Chinese, with exercises. Tucson: University of Arizona Press.

Tierney, R.T., & Mosenthal, J.H. (1983). Cohesion and textual coherence. Research in the teaching of English, 17, 215-229.

Vande Kopple, W.J. (1982). Functional sentence perspective, composition, and reading. College composition and communication, 30, 50-63.

Vande Kopple, W.J. (1983). Something old, something new: Functional sentence perspective. Research in the teaching of English, 17, 85-99.

Vachek, J. (1966). The lingustic school of Prague: An introduction to its theory and practice. Bloomington: Indiana University Press.

Winterowd, W.R. (1970). The grammar of coherence. College English, 31, 328-35.

Witte, S.P. (1983). Topical structure and revision: An exploratory study. College composition and communication, 33, 313-341.

Witte, S.P., & Faigley, L. (1981). Coherence, cohesion, and writing quality. College composition and communication, 32, 189-204.

## Appendix

1. The psychology passage used in this study is the entire introductory section of M. Sibicky & J.F. Dovidio's "Stigma of psychological therapy: stereotypes, interpersonal reactions, and self-fulfilling prophecy," Journal of Counseling Psychology 33 (1986): 148-54. [Copyright 1986 by the American Psychological Association. Reprinted by permission.] The text, with parenthetical bibliographical references deleted, is as follows. The paragraphs and sentences have been numbered, and a double slash has been placed between the theme and rheme of each sentence.

1.1 Research has demonstrated that // first impressions and stereotypes can influence social interactions in ways that lead to their behavioral confirmation--even to the extent of causing mistaken impressions to become real. 1.2 In one study, for example, // Snyder, Tanke, and Berscheid investigated the process of behavioral confirmation of the stereotype associated with physical attractiveness. 1.3 Their results revealed that // men formed more favorable first impressions of female targets when they were led to believe that the target was physically attractive than when they thought she was unattractive. 1.4 Consistent with these first impressions, // women interacting with men who believed that they were attractive then came to behave in a manner that was more socially desirable than did women conversing with partners who believed that they were unattractive. 1.5 A self-fulfilling prophecy // thus occurred. 1.6 Snyder et al. concluded that // male perceivers used different styles of interaction for the two groups of targets. 1.7 These behaviors, in turn, // guided and constricted the behavioral options of female targets in ways that led them to conform to men's initial impressions.

2.1 The reasoning outlined by Snyder et al. // is similar in many ways to the processes proposed by Becker and Scheff relating to the labeling approach to social deviance. 2.2 Labeling theory // suggests that once a person is labeled as mentally ill, preexisting stereotypes are activated in other people. 2.3 With respect to mental illness, // the public perceives mental patients and ex-mental patients as threatening and socially undesirable. 2.4 According to labeling theory, // based on these perceptions, people systematically alter their expectations, vocabulary, and response cues when they interact with mental patients or ex-mental patients. 2.5 Individuals who are labeled as mentally ill // may then incorporate others' expectations into their own self-concepts, thereby leading to further loss of self-control and continued deviant behavior.

3.1 Although there has been little systematic investigation, // it appears that seeking psychological therapy may be associated with stigmatization similar to that associated with

being mentally ill. 3.2Describing a person as seeking psychological therapy, like labeling a person as mentally ill, // implies that the person has psychological problems and is incapable of handling his or her own problems. 3.3Phillips, for example, // found that although the most negative attitudes were expressed toward people who had been in a mental institution, negative attitudes were also displayed toward people who sought psychological assistance from a clergyman or from a psychiatrist. 3.4Also, research by Goodyear and Parish and by Parish and Kappes // indicates that a person described as seeking counseling is rated more negatively than is a "typical" person. 3.5Thus, a possible consequence for seeking psychological aid // may be negative evaluations and rejection from others.

4.1The present experiment // investigated whether negative social perceptions currently exist concerning persons who seek psychological therapy at a University Counseling Center and examined how these negative social perceptions, if they exist, influence actual dyadic social interactions. 4.2Piner and Kahle // demonstrated that stigmatization toward people labeled as mental patients is particularly strong in personal, ego-involving situations. 4.3Using a procedure similar to that used by Snyder et al., // previously unacquainted subjects interacted in a conversation that had been structured to control the information that one randomly chosen dyad member (designated as the perceiver) received concerning the client status of the other dyad member (designated the target). 4.4Client status information // was controlled by the experimenter, who either informed the perceiver that the dyad partner was currently in psychological therapy at the University Counseling Center (client) or made no mention of the target as a client at the counseling center (nonclient). 4.5The target // was unaware of the experimental manipulation. 4.6This experimental procedure // provided an opportunity to test directly several predictions derived from labeling theory. 4.7It was hypothesized that // the perceiver would form more negative impressions of the partner when the target was believed to be seeking psychological aid than when no mention was made about the target seeking therapy. 4.8Furthermore, it was predicted that // the perceiver would interact with the target in ways that would lead to changes in the target's behavior that would confirm the perceiver's initial impressions. 4.9Consequently, targets who were believed to be seeking psychological aid // were expected to behave in a less socially desirable manner as compared to targets who were not believed to be seeking therapy.

2. The biology passage is the entire introductory section of B.S. Cohen, N.H. Harley, & T.C. Tso's "Clearance of polonium-210-enriched cigarette smoke from the rat trachea and lung," Toxicology and Applied Pharmacology 79 (1985): 314-322. The text, with parenthetical bibliographical references deleted, is as follows:

1.1 Polonium-210 and lead-210, members of the natural uranium series, // are found in cigarette tobacco. 1.2 The transfer of $^{210}Po$ to the smoke and its presence in samples of bronchial epithelium from cigarette smokers // were documented by Radford and Hunt. 1.3 When a cigarette is smoked, // about 10% of the $^{210}Po$ is transferred to the mainstream smoke. 1.4 The concentration in mainstream smoke // is about 0.5 pCi/g of smoke tar. 1.5 Some of the [alpha] activity // deposits in the lung and may be involved as a causative agent in lung cancer. 1.6 Ninety percent of the lung cancers in cigarette smokers // are bronchogenic. 1.7 Thus, the significant [alpha] radioactivity // will be the amount on the airway surfaces. 1.8 The average concentration on the bronchial surface // is expected to be very low based on knowledge of aerosol deposition and clearance. 1.9 A few measurements // have been made of the activity on the mucosal surface in human autopsy specimens 1.9 and the average activity // is a few 10ths of a $fCi/cm^2$. 1.10 In about one of three smokers, // hot spots, or areas of accumulation, have been found. 1.11 These // could deliver a radiation dose to the sensitive basal cells of the bronchial epithelium sufficient to induce cancer, depending upon residence time. 1.12 So far, measurements // have only been made at one point in time, 1.12 and it is difficult // to infer cumulative [alpha] dose from the data.

2.1 The work reported here // was designed primarily to visualize the spatial and temporal distribution of activity on the trachea and major bronchi of rats exposed to smoke from cigarettes made with tobacco grown to incorporate a highly elevated content of $^{210}Pb$ in the leaf. 2.2 The lung burden of $^{210}Po$ was also measured as a function of time // because it is of considerable interest to compare the measurements with results predicted by a model of the deposition of $^{210}Pb$ and $^{210}Po$ in the trachea and pulmonary parenchyma. 2.3 Agreement // indicates that the mechanisms which govern the deposition and removal of submicron particles are reasonably well predicted by present theoretical concepts.

3.1 Most of the $^{210}Po$ in cigarette smoke // is associated with insoluble particles. 3.2 This study // is concerned with the clearance of the insoluble particles from the pulmonary and tracheo-bronchial regions. 3.3 The soluble $^{210}Pb$ // tanslocates mainly to bone 3.3 and $^{210}Po$ // distributes throughout the body tissues.

3. The history passage is the entire introductory section of T.F. Remington's "Politics and professionalism in Soviet journalism," Slavic Review 44 (1985): 489-503. The text, sans footnotes, is as follows:

[1.1]The Western ideal of jornalistic objectivity, influenced by liberal principles of the rulers' accountability to the ruled and the empirical skepticism of science, // developed as an occupational response by journalists to marketplace competition among commercially or politically motivated suppliers of information and came to define the journalist's professional ethic of impartiality and independence. [1.2]At the same time the term professionalism // must be used advisedly. [1.3]Journalism // is a field with some but not all of the attributes of a profession. [1.4]By the usual tests of the freedom of the practitioners to govern entry and exit from the field, to possess an exclusive right to carry on their trade, and to set the standards of performance, // journalists are not as autonomous as, for example, physicians and attorneys. [1.5]If they sought to close the shop to outsiders or to set standards of writing and reporting, // they would be infringing upon the prerogatives of "management"--editors and publishers. [1.6]Nevertheless, the need to define an occupational role by reference to values of objectivity and impartiality, mastery of technique, and freedom from outside interference // creates strong pressures for professionalism.

[2.1]At the same time, the literature on the sociology of news // has shown that more immediate influences than these general principles affect the results of the journalist's labor. [2.2]Above all the recent literature // has emphasized the importance of organizational factors in shaping the day-to-day norms of journalistic and editorial behavior. [2.3]Media organizations // adopt standardized routines in selecting and presenting information that reflect the influence of their own internal economies based on the relative scarcities of time, staff, air time, or column space. [2.4]Adaptation to the needs of the organization // produces characteristic "structural" bias in news reporting, for instance the "headline" approach to information, concentration of attention on a small number of organizations and individuals that regurlarly produce reportable events or utterances, and reportage which shows to best advantage the peculiarities of a given medium. [2.5]These traits // illustrate the point that marketplace competition and the ideal of objectivity influence the journalistic product not directly but through the internal laws of the media organization.

[3.1]In the Soviet Union, where journalistic practice must in principle reflect party-mindedness, ideological orthodoxy, and political loyalty, // how thoroughly are the working rules of journalism pervaded by the general principles laid down by party doctrine? [3.2]Do the day-to-day requirements of the job // generate tendencies toward a code of journalistic professionalism

that might, given time and favorable circumstances, spur journalists to form an institutional counterweight to party authority? 3.3 One of the standing phrases party officals use // is that the media are "an integral part of the party apparatus." 3.4 In this article // I shall consider whether and by what means the party works to realize this aim. 3.5 Four aspects of journalism bearing on these questions // will be examined: the nature of journalistic training in school, the lessons that early exposure to journalistic practice teaches, the role of the Journalists' Union, and the social standing of journalism in Soviet society.